A Proposal For:

BEFORE-AFTER CRASH ANALYSIS:
A PRIMER FOR USING THE EMPIRICAL BAYES METHOD

Submitted by

JODI L. CARSON, PH.D.
ASSISTANT PROFESSOR
DEPARTMENT OF CIVIL ENGINEERING
MONTANA STATE UNIVERSITY
214 COBLEIGH HALL
BOZEMAN, MONTANA 59717

Submitted to

MONTANA DEPARTMENT OF TRANSPORTATION
RESEARCH MANAGEMENT UNIT
2701 PROSPECT AVENUE
HELENA, MT 59620

September 12, 2003

TABLE OF CONTENTS

1 PROBLEM STATEMENT 1
1.1 BACKGROUND SUMMARY 2
1.1.1 Before/After Analyses 2
1.1.2 Case/Control Analyses 3
1.1.3 Regression Methods 3
1.1.4 Empirical Bayes Method 4
1.2 OBJECTIVES 5
1.3 BENEFITS 5
2 RESEARCH PLAN 6
2.1 WORK TASKS 6
2.1.1 Work Task 1 – Statistically Confirm Previous Crash Trend Findings 6
2.1.2 Work Task 2 – Develop Draft and Final Report 6
2.1.3 Work Task 3 – Develop Empirical Bayes Method Primer 7
2.1.4 Work Task 4 – Conduct Empirical Bayes Method On-Site Training Session 7
2.2 PRODUCTS 8
2.3 IMPLEMENTATION 8
3 TIME SCHEDULE 9
4 PROJECT BUDGET 10

LIST OF TABLES

Table 1. Crash Trends for Interstate Pavement Preservation Projects (MDT 2002) 1
Table 2. Project Schedule 9

1 PROBLEM STATEMENT

A significant portion of the Montana Department of Transportation's (MDT) resources is spent on reconstruction and pavement preservation projects. While the costs for these projects are well quantified, the resulting benefits are not, particularly with respect to traffic safety.

MDT's Strategic Business Plan calls for "providing a safe and efficient intermodal transportation system." In an effort to better determine the safety-related benefits attributable to reconstruction and pavement preservation activities, MDT's Safety Management Section performed a before/after, case/control analysis that considered crash trends at various sites around the state. Specifically, "case" sites comprising reconstruction and pavement preservation projects completed around the state in late 1996, 1997 and 1998 were obtained from the Construction Bureau. Comparable "control" sites with similar roadway geometry and traffic characteristics were next identified. Crash data were researched for three years prior to and three years following construction activity. Categorized by roadway functional class and project type (i.e., Interstate, non-Interstate pavement preservation, non-Interstate reconstruction and urban reconstruction), the study compared crash frequency, crash rate, severity and severity index: (1) prior to and following construction activity, (2) between the case and control sections and (3) between the case section and statewide averages. An example of the findings is provided in Table 1 for Interstate pavement preservation projects.

Table 1. Crash Trends for Interstate Pavement Preservation Projects (MDT 2002)

Crash Rate	Case	Control	Statewide Average
Before	0.98	0.90	1.09
After	0.92	0.85	1.04

Severity	Case	Control	Statewide Average
Before	2.33	2.27	2.29
After	1.87	2.05	2.03

"In conclusion, the before/after construction crash rate change did not improve compared to the control group and the statewide average, however crash severity for interstate projects has improved." (MDT 2002)

While this initial study suffered case/control section similarity and temporal challenges common to before/after analyses (e.g., changes in traffic stream composition, automobile design, statutory speed limits, property damage reporting thresholds, etc.), its primary limitation, and hence, the motivation for this second study, relates to the absence of any statistical confirmation of these findings. Identified by MDT's Safety Management Section personnel as a next step is to "have a statistician evaluate these data and establish the statistical framework for future evaluations." This study would accomplish just that.

1.1 BACKGROUND SUMMARY

Accurate estimation of safety-related benefits attributable to overall construction activities (i.e., reconstruction or pavement preservation projects) or, as more commonly investigated, attributable to a specific safety treatment, is critical to avoid over-estimation of related benefits, inappropriate inferences regarding resulting benefits and/or widespread use of a potentially ineffective safety treatment.

A brief methodological review suggests four predominant methods used in practice to evaluate safety-related benefits: (1) before/after analyses, (2) case/control analyses, (3) regression methods and (4) the Empirical Bayes method. In some instances, a combination of methods is applied to overcome individual method shortcomings.

1.1.1 Before/After Analyses

Simple before/after analyses estimate the safety-related benefits of an improvement by most commonly comparing crash occurrence at a location before and after some "treatment" (in this context, the term "treatment" will include construction-related activities as well as other types of targeted safety improvements). The comparison can best be expressed with a reduction factor, which provides the percentage of the original crashes that is prevented by the treatment:

$$?_i \; ? \; \frac{(N_{B_i} / V_{B_i}) \; ? \; (N_{A_i} / V_{A_i})}{(N_{B_i} / V_{B_i})}$$

where

$?_i$ denotes the crash reduction factor at site i for the specific crash type,

N_{Bi} and N_{Ai} are the number of crashes of that type at site i before and after the improvement, respectively and

V_{Bi} and V_{Ai} represent the traffic entering site i in millions of vehicles for the before and after period, respectively.

Simple before/after analyses are susceptible to temporal variations (i.e., adverse weather-related trends, changes in traffic volumes and traffic stream composition, regulatory changes, etc.) and as such may lead to inaccurate or exaggerated inferences regarding safety-related treatment effects. A noted increase or decrease in crashes may result from the random nature of crash occurrence independent of any treatment. This phenomenon is particularly problematic when post-treatment data is limited to only a few years.

1.1.2 Case/Control Analyses

To overcome the shortcomings of the simple before/after analysis, temporal observations are often combined with observations made across "case" and "control" sites. A group of comparison sites (i.e., control) with geometric and site characteristics similar to the site being studied (i.e., case) is identified. Ideally, the control sites should have roadway geometrics, traffic volumes and land use characteristics identical to that of the study site. Crash data is collected for the same before/after time period at both the case and control sites.

The inclusion of control sites in before/after analyses helps to ensure that any observed change in crash occurrence is attributable to the treatment and not confounding factors or systemic changes (i.e., if the same reduction in crashes was noted before and after at the control site without any treatment and the case site with the treatment, the treatment can be presumed ineffective).

The challenge in conducting combined before/after, case/control analyses is identifying a sufficient number and quality of comparison sites. Oftentimes, either a smaller number of highly-similar sites or a larger number of less similar sites is the resulting compromise.

1.1.3 Regression Methods

Rather than trying to control for the varying roadway geometrics, traffic volumes and land use characteristics through control site selection, regression methods can be used to directly account for these factors and their affect on safety. Most widely applied and easily understood, the multiple linear regression equation assumes the following form:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

where:

Y denotes the dependent variable (e.g., crash rate)

X_{i1} through $X_{i,p-1}$ denote independent, explanatory variables (e.g., lane width, degree of horizontal curvature, average annual daily traffic, percent trucks in the traffic stream, etc.)

β_0 through β_{p-1} denote estimable parameters, β_0 represents the y-intercept value

ε_i is the unexplainable, random error

More simplistic linear regression models that consider only crash exposure as an explanatory variable have also been employed however, a significant body of research now points to the appropriate use of more sophisticated regression techniques including Poisson, negative binomial and zero-inflated regression models to estimate crash occurrence and logistic and ordered probit regression to estimate crash severity. Application of these advanced modeling techniques is limited in practice because of the large amount of data required and the complexity of the analysis.

Further, since regression methods cannot take into account every factor that influences the occurrence or severity of a crash, estimates of resulting safety benefits may suffer similar accuracy problems as other methods. Mountain, Fawaz and Jarrett (1996) observed that the use of regression methods to predict safety benefits attributable to minor intersection treatments underestimated potential benefits by 30 to 40 percent.

1.1.4 Empirical Bayes Method

Combining the strengths of the before/after, case/control and regression methods, the Empirical Bayes (EB) method for estimating safety-related benefits has emerged as the preferred method in practice.

The EB method uses data from a group of similar control sites as well as pre-treatment data from the case site to estimate how many crashes would have occurred at the study site had no improvements (i.e., treatment) been made. This allows for the comparison of the "after" treatment crash rate at the site to the estimated or expected crash rate at the site in the absence of treatment during the same time period.

The Empirical Bayes model for the expected number of crashes, k, given that K crashes have already been observed at a particular site is as follows:

$$E\{k/K\} = ?\, E\{k\} + (1-?)K$$

where ? is a coefficient between 0 and 1 calculated as:

$$? \; ? \; \frac{1}{1 \; ? \; ?\, VAR\{k\} / E\{k\} \, ?}$$

where VAR{k} is the variance of the number of crashes for the reference population k. In practice, E{k} and VAR{k} are estimated using either multivariate regression techniques or the method of sample moments.

The EB method increases the precision (reduces the variability) of estimates beyond what is possible using any of the above methods, particularly when limited to the use of a two- to three-year crash history. The theory of the EB method is well developed and significant efforts are currently underway at the national level to integrate this theory into practice. It is now used in the USDOT's *Interactive, Highway Safety Design Model (IHSDM)* and will be used in their *SafetyAnalyst* software forthcoming in 2006 making this method the standard in professional practice.

A second benefit of the EB method, though not of direct concern in this investigation, is its ability to account for regression-to-the-mean bias. When crash rates are used as a heavily-weighted determinant in safety treatment site selection, it follows that selected sites are likely to have greater than average number of crashes during the pre-improvement period (i.e., sites selected for treatment are more likely to have high crash rates during the years prior to treatment because this motivated their selection). A noted reduction in crashes at these sites may result independent of faulty design or unsafe conditions but may instead be a result of the random nature of crash occurrence; a high crash rate will likely decrease over time - with or without treatment. When sites have been selected because of a high crash rate during the period prior to the treatment and this "before" rate is used to estimate the reduction in crash occurrence, the random variation in crashes will lead to exaggerated estimates of the treatment's effectiveness (Hauer 1986).

Because safety is not a primary determinant in site selection for reconstruction or pavement preservation projects, regression-to-the-mean bias is not of concern in this investigation. However, more widespread application of the EB method by MDT Safety Management Section personnel following this investigation will result in more accurate and robust inferences regarding safety treatment effects.

1.2 OBJECTIVES

The overall goal of this study is to facilitate the transfer of Empirical Bayes statistical theory into practice for MDT Safety Management Section personnel. Specifically, this project will accomplish the following objectives:

?? statistically confirm the findings reported in the initial before/after, case control analysis (MDT 2002) using the Empirical Bayes method and simultaneously provide proof-of-method and

?? effectively implement these methods in MDT's Safety Management Section though: (1) development of a Primer that effectively demonstrates use of the Empirical Bayes method for analyzing data and (2) an on-site training session for MDT Safety Management Section personnel.

1.3 BENEFITS

Most directly and immediately, this study will provide statistically valid conclusions related to the safety benefits attributable to a subset of reconstruction and pavement preservation projects completed in 1996, 1997 and 1998.

This study will also result in an enhanced skill set for MDT Safety Management Section personnel that will allow them to monitor crash trends over time related to reconstruction and pavement preservation projects using the same statistically robust methods. This enhanced skill set can also be applied to a wider variety of investigations commonly performed by MDT Safety Management Section personnel (i.e., the effect of specific safety treatments), improving the overall understanding of traffic safety.

Ultimately, an improved understanding of traffic safety would result in an improved overall level of safety in the state of Montana, providing a significant cost savings to both MDT and the motoring public.

2 RESEARCH PLAN

This study will be completed by Dr. Jodi Carson, Assistant Professor in the Department of Civil Engineering at Montana State University. Dr. Carson's technical skills lie in the area of advanced statistical applications, including biostatistics and econometrics, to transportation challenges. She has applied this core technical skill to a diverse set of topics related to Intelligent Transportation Systems, safety and incident management, trucking and commercial vehicle operations, human factors and, to a lesser extent, public transit during her 11-year career.

2.1 WORK TASKS

The work plan for this project consists of four tasks that are described more fully below.

2.1.1 Work Task 1 – Statistically Confirm Previous Crash Trend Findings

In December 2002, MDT's Safety Management Section published a report that considered crash trends at various sites around the state an effort to better determine the safety-related benefits attributable to reconstruction and pavement preservation activities. Specifically, "case" sites comprising reconstruction and pavement preservation projects completed around the state in 1996, 1997 and 1998 were compared to "control" sites with similar roadway geometry and traffic characteristics for three years before and after the project completions. Qualitative findings related to changes in crash frequency, crash rate, severity and severity index: (1) prior to and following construction activity, (2) between the case and control sections and (3) between the case section and statewide averages were reported separately for Interstate, non-Interstate pavement preservation, non-Interstate reconstruction and urban reconstruction projects.

Utilizing this same set of data, researchers will perform an Empirical Bayes analysis to statistically confirm the qualitative trends observed in the initial study. This exercise will not only provide valuable information pertaining to the safety benefits attributable to reconstruction and pavement preservation projects but will also provide proof-of-method for this type of application and the basis for demonstrative examples to be included in the Empirical Bayes Method Primer (described later in this document). The method used to determine the expected number of crashes at a site, E{k}, and their variability, VAR{k}, (i.e., multivariate regression techniques or the method of sample moments) will be governed by data availability.

2.1.2 Work Task 2 – Develop Draft and Final Report

Because the findings from Work Task 1 (i.e., the safety benefits attributable to reconstruction and pavement preservation projects) are of interest in themselves, a stand-alone Draft and Final Report will be produced. This report will provide a brief description of the Empirical Bayes method and a summary and interpretation of the analysis results separately for Interstate, non-Interstate pavement preservation and non-Interstate reconstruction projects (data limitations preclude statistical analysis of urban reconstruction projects). Limitations to the data, the analysis and the subsequent results will be discussed to prevent inappropriate transfer or extrapolation of these findings.

2.1.3 Work Task 3 – Develop Empirical Bayes Method Primer

Following successful application to and reporting of the Empirical Bayes method using MDT crash data, a detailed, step-by-step Empirical Bayes Method Primer for use by MDT Safety Management Section personnel will be developed. This Primer will focus on the application and interpretation the EB method for investigating safety effects. Specifically, the Primer will address the following major topics:

?? supporting data needs and recommended format for EB analysis

?? EB methodology, focused on application and procedure rather than theory and

?? appropriate interpretation and reporting of findings.

The Primer will discuss both methods used to determine the expected number of crashes at a site, E{k}, and their variability, VAR{k}, (i.e., multivariate regression techniques or the method of sample moments) though the numerical examples drawn from Work Task 1 may only illustrate one method.

Secondary topics to be covered include the following:

?? variations to the EB method, primarily driven by data limitations

?? other appropriate applications of the EB method and conversely, applications when the EB method should not be applied and

?? additional references for further information.

Numerical examples will be drawn from Work Task 1 and provided throughout the Primer for illustrative purposes. Software applications (e.g., Excel worksheets) that ease the analysis or reporting process will be integrated with and included in the Primer.

Special attention will be paid to ensure that the Primer is written in an easy-to-understand language, easy to follow and sufficiently broad to address the variety of applications that may be needed by MDT Safety Management Section personnel.

2.1.4 Work Task 4 – Conduct Empirical Bayes Method On-Site Training Session

Upon successful review and revision of the Empirical Bayes Primer, an on-site training session will be scheduled and facilitated in Helena, Montana at the Montana Department of Transportation Headquarters Building. Attendees will include personnel from MDT's Safety Management Section who are currently or may in the future be responsible for performing analyses utilizing the Empirical Bayes method. Ideally, decision-makers responsible for directing policy and allocating resources based on findings from the Empirical Bayes method would be in attendance as well. A four-hour training session is anticipated.

2.2 PRODUCTS

This project will result in the following products:

- a Draft and Final Report summarizing the statistically-confirmed effects of reconstruction and pavement preservation projects completed in 1996, 1997 and 1998 on traffic safety
- a Primer for use by MDT Safety Management Section personnel focused on the application and interpretation the Empirical Bayes method for investigating safety effects and
- an on-site training session in Helena, Montana for MDT Safety Management Section personnel to effectively introduce the Empirical Bayes method and the supporting Primer

Because of the short duration of project (five months), quarterly progress reports will not be submitted. Each of these products will be thoroughly reviewed by Sue Sillick, Research Bureau Chief and appropriate Safety Management Section personnel prior to completion.

2.3 IMPLEMENTATION

The findings from this study will facilitate immediate and long-term implementation. This study will provide statistically valid conclusions related to the safety benefits attributable to a subset of reconstruction and pavement preservation projects completed in 1996, 1997 and 1998 that is of value to decision-makers.

This study will provide MDT Safety Management Section personnel with an enhanced skill set that will allow them to monitor crash trends over time related to reconstruction and pavement preservation projects using the same statistically robust methods. This enhanced skill set can also be applied to a wider variety of investigations commonly performed by MDT Safety Management Section personnel (i.e., the effect of specific safety treatments), improving the overall understanding of traffic safety.

3 TIME SCHEDULE

The estimated project completion schedule is depicted in Table 2. The estimated start date of this project is September 1, 2003 and the estimated duration is five months resulting in a completion date of January 31, 2004.

Table 2: Project Schedule

		2003				2004
WORK TASKS		SEP	OCT	NOV	DEC	JAN
1	Statistically Confirm Previous Crash Trend Findings			MDT Review		
2	Develop Draft and Final Report					MDT Review
3	Develop Empirical Bayes Method Primer					
4	Conduct Empirical Bayes Method On-Site Training Session					

4 PROJECT BUDGET

Project Title:	Before-After Crash Analysis: A Primer for using the Empirical Bayes Method		
Period of Project:	September 1, 2003 - January 31, 2004		
Principle Investigator:	Jodi L. Carson, Ph.D.		
Direct Labor	**Est. Hours**	**Rate/Hour**	**Est. Cost**
Jodi Carson	174.00	$ 36.00	$ 6,264.00
Graduate Students		$ 12.07	$ -
Undergraduate Students		$ 9.50	$ -
Business Manager	8.00	$ 15.83	$ 126.64
Administrative Support		$ 10.54	$ -
		Total:	$ 6,390.64
		Benefits @ 25%:	$ 1,597.66
		Total Direct Labor:	$ 7,988.30
Supplies/Expendables			**Est. Cost**
Copies, paper, etc.			$ 100.00
		Total Supplies:	$ 100.00
Communications			**Est. Cost**
Telephone, postage, etc.			$ 50.00
		Total Communications:	$ 50.00
Travel			**Est. Cost**
In-state travel (to/from Helena)			$ 100.00
Out-of-state travel			
		Total Travel:	$ 100.00
		TOTAL DIRECT COST:	$ 8,238.30
		Indirect Cost Rate (20%):	$ 1,647.66
		TOTAL PROJECT COST:	$ 9,885.96

REFERENCES

Hauer, Ezra. On the Estimation of the Expected Number of Accidents. Accident Analysis and Prevention. Volume 18, Number 1. February 1986. pp. 1-12.

Mountain, L., B. Fawaz and D. Jarrett. Accident Prediction Models for Roads with Minor Junctions. Accident Analysis and Prevention. Volume 28. Number 6. November 1996. pp. 695-707.